Assessment of the InSiGHT Interpretation Criteria for the Clinical Classification of 24 MLH1 and MSH2 Gene Variants

Assessment of the InSiGHT Interpretation Criteria for the Clinical Classification of 24 *MLH1* and *MSH2* Gene Variants

Rossella Tricarico[1,2], Mariann Kasela[3], Cristina Mareni[4], Bryony A. Thompson[5,6], Aurélie Drouet[7], Lucia Staderini[1,8], Greta Gorelli[1], Francesca Crucianelli[1], Valentina Ingrosso[1], Jukka Kantelinen[3], Laura Papi[1], Maria De Angioletti[9], Margherita Berardi[9], Pascaline Gaildrat[7], Omar Soukarieh[7] Daniela Turchetti[10], Alexandra Martins[7], Amanda B. Spurdle[11], Minna Nyström[3], Maurizio Genuardi[1,12]*, and InSiGHT Variant Intepretation Committee[13]

[1]Department of Biomedical, Experimental and Clinical Sciences, Medical Genetics Unit, University of Florence, Florence, Italy

[2]Current address: Cancer Epigenetics and Cancer Biology Programs, Fox Chase Cancer Center, Philadelphia, United States

[3]Department of Biosciences, Division of Genetics, University of Helsinki, Helsinki, Finland

[4]Tuscan Tumor Institute, Florence, Italy

[5]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, United States

[6]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Victoria, Australia

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/humu.23117.

[7]Inserm-U1079-IRIB, University of Rouen, Normandy Centre for Genomic and Personalized Medicine, Rouen, France

[8]Current address: Microbiology and Virology Department, Fondazione IRCCS Policlinico San Matteo, Pavia

[9]Cancer Genetics and Gene Transfer - Core Research Laboratory, Istituto Toscano Tumori, Florence, Italy

[10]Medical Genetics; Department of Medical and Surgical Sciences (DIMEC); University of Bologna, Bologna, Italy

[11]Genetics and Computational Biology Department, QIMR Berghofer Medical Research Institute, Brisbane, Australia

[12]Current affiliations: Institute of Genomic Medicine, "A. Gemelli" School of Medicine, Catholic University of the Sacred Heart; and Medical Genetics Unit, Fondazione Policlinico Universitario "A. Gemelli", Rome, Italy

[13]InSiGHT: International Society for Gastrointestinal Hereditary Tumors; for a full list of InSiGHT contributors see footnote after the Acknowledgments section

*Corresponding author:

E-mail address: maurizio.genuardi@unicatt.it

Short title: Interpretation of MMR Gene Variants

Key words: Lynch Syndrome; Functional assays; Splicing; Variants of Uncertain Significance (VUS); Multifactorial analysis; Microsatellite instability.

ABSTRACT

Pathogenicity assessment of DNA variants in disease genes to explain their clinical consequences is an integral component of diagnostic molecular testing. The International Society for Gastrointestinal Hereditary Tumors (InSiGHT) has developed specific criteria for the interpretation of mismatch repair (MMR) gene variants. Here, we performed a systematic investigation of 24 *MLH1* and *MSH2* variants. The assessments were done by analyzing population frequency, segregation, tumor molecular characteristics, RNA effects, protein expression levels and *in vitro* MMR activity. Classifications were confirmed for 15 variants and changed for 3, and for the first time determined for 6 novel variants. Overall, based on our results we propose the introduction of some refinements to the InSiGHT classification rules. The proposed changes have the advantage of homogenizing the InSIGHT interpretation criteria with those set out by the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) consortium for the *BRCA1/BRCA2* genes. We also observed that the addition of only few clinical data was sufficient to obtain a more stable classification for variants considered as "likely pathogenic" or "likely non pathogenic". This shows the importance of obtaining as many as possible points of evidence for variant interpretation, especially from the clinical setting.

INTRODUCTION

Lynch syndrome (LS) (MIM# 120435) is the most common form of inherited colorectal and endometrial cancer, predisposing also to other gastrointestinal (GI) (stomach, small bowel, biliary tract, pancreas) and non-GI cancers (urinary tract, ovary and others). The syndrome is transmitted as an autosomal dominant trait, and caused by constitutional defects in the mismatch repair (MMR) genes *MLH1* (MIM# 120436), *MSH2* (MIM# 609309) *MSH6* (MIM# 600678) and *PMS2* (MIM# 600259) [Lucci-Cordisco et al., 2003; Lynch & De la Chapelle, 2003]. Detection of a constitutional loss-of-function variant in an MMR gene provides diagnostic confirmation of LS and is essential for the identification of at-risk members in LS families through predictive testing. This is especially important for LS care, since carriers of MMR gene pathogenic variants can benefit from different risk reducing options, including stringent surveillance protocols, prophylactic surgery, and chemoprevention [Vasen et al., 2013].

A major challenge in the diagnosis and management of LS is the frequent occurrence of variants of uncertain significance (VUS) in the MMR genes. Depending on the gene, about $1/5^{th}$ to $1/3^{rd}$ of DNA sequence variants identified during the course of LS clinical testing are of uncertain significance [Sijmons et al., 2013], limiting risk reduction and management options in probands and preventing their use in predictive gene testing in relatives.

The classification of DNA sequence variants identified in MMR and other cancer predisposition genes is recommended to be based on data from multiple sources, including clinical observations, tumor pathology studies and several RNA and protein based functional assays [Couch et al., 2008; Hofstra et al., 2008; Spurdle et al., 2008; Tavtigian et al., 2008a; Richards et al., 2015]. A number of *in silico*

programs have also been devised to assist with the prediction of functional consequences of inherited MMR gene alterations [Tavtigian et al., 2008b; Thompson et al., 2013b; Niroula & Vihinen, 2015].

Recently, the International Society for Gastrointestinal Hereditary Tumors (InSiGHT) has developed criteria for the interpretation of MMR gene variants, with the aim to improve the clinical utility of genetic testing for LS. A systematic clinical classification of all variants contained in InSiGHT locus specific databases (http://insight-group.org/variants/database/) was performed, based on a multifactorial bayesian quantitative approach and/or on stringent combinations of qualitative clinical and functional data [Thompson et al., 2014]. Variants were classified using a 5-tier system devised for cancer predisposing genes [Plon et al., 2008].

In the present study, we assessed the pathogenicity of 24 MMR gene variants using an extensive combination of RNA and protein-based functional assays, segregation studies, and tumor analyses. We were able to classify 6 novel variants as well as to confirm or refine the classification of 18 variants previously assessed by InSiGHT. Overall, we show the necessity of using different analyses in VUS classification and discuss their specific value and status in the interpretation process.

MATERIALS AND METHODS

Genetic variants, patients and samples

The variants assessed in this study were detected in a single laboratory in families fulfilling the

Bethesda criteria [Vasen, 2005] ascertained through cancer family clinics from 2002 to 2011. Overall, 57 MMR gene variants (25 *MLH1* and 32 *MSH2*), excluding well established polymorphisms, were detected in a total of 56 families. Variants that were clearly disease causing (ie truncating, splicing alterations, or large rearrangements), as well as established polymorphisms, were not considered. Exceptions were some alleles that, despite being currently considered polymorphic, were not clearly classified at the time of their detection; these included *MLH1* c.1558+14G>A, and *MSH2* c.380A>G and c.1511-9A>T, whose assessment was useful for the validation of the variables investigated for classification. Overall, we evaluated 24 MMR gene variants, 13 *MLH1* and 11 *MSH2*, identified in 37 unrelated families. All variants were single nucleotide substitutions at the genomic level. Based on their positions in the DNA sequence and predicted effects, they could be divided into the following groups: 12 aminoacid substitutions (8 in *MLH1* and 4 in *MSH2*), 1 *MLH1* potential splice site change affecting the first exonic position and also potentially causing an aminoacid substitution, 4 synonymous exonic nucleotide substitutions (1 *MLH1*, 3 *MSH2*), and 7 intronic changes outside the most conserved positions of the splice site consensus sequences (4 *MLH1* and 3 *MSH2*). Six variants (4 *MLH1* and 2 *MSH2*) were previously unreported and therefore not assessed by InSiGHT.

Peripheral blood leukocyte (PBL) samples were collected from 76 and 16 subjects for DNA and RNA extractions, respectively. Furthermore, lymphoblastoid cell lines (LCLs) were established from 7 variant carriers. Forty-nine paraffin-embedded tumor specimens were obtained from 42 carriers of 23 different variants.

The study was approved by the Institutional Ethical Board of the Careggi University Hospital, Florence. Informed consent was obtained from all patients for the use of specimens and clinical/pathological data for research purposes.

Molecular analyses

The complete coding sequence and flanking exon–intron borders of the *MLH1* and *MSH2* genes were investigated by direct sequencing on genomic DNA. The presence of *MLH1* and *MSH2*, as well as of *EPCAM* (MIM# 185535), genomic rearrangements and the methylation status of the *MLH1* promoter were investigated by Multiplex Ligation-dependent Probe Amplification (MLPA), as previously described [Crucianelli et al., 2014]. Values lower than 0.15 were assumed as a cut-off for normal methylation levels according to previous studies [Gylling A et al., 2007; Crucianelli et al., 2014].

The identified *MLH1* and *MSH2* variants have been submitted to the InSiGHT MMR gene variant database (http://www.insight-group.org/variants/database/). Variants were defined according to the Human Genome Variation Society (HGVS) recommendations [den Dunnen et al., 2016]. DNA variant numbering is based on the *MLH1* and *MSH2* cDNA sequences (GenBank accession numbers NM_000249.2 and NM_000251.1, respectively) with the A of the ATG translation–initiation codon numbered as +1. Aminoacid numbering starts with the translation initiator methionine as +1.

To investigate the presence of the p.Val600Glu (V600E) hotspot mutation, *BRAF* (MIM# 164757) exon 15 was directly sequenced in 7 tumor samples of *MLH1* variant carriers, using previously reported primers and conditions [Mancini et al., 2010].

Loss of heterozygosity (LOH) analysis of the regions containing the identified variants was performed on matched leukocyte and tumor tissues from 19 probands by analysis of direct sequencing electropherograms [Janssen et al., 2011; Janssen et al., 2012].

Microsatellite instability and immunohistochemical analyses

A total of 47 tumor samples and matched normal mucosa or PBLs from 40 patients were evaluated for microsatellite instability (MSI) using a 5-mononucleotide marker panel [Suraweera et al., 2002; Buhard et al., 2006; Giunti et al., 2009]. Tumors were classified into three categories according to the proportions of markers showing instability: MSI-H (high-level MSI), MSI-L (low-level MSI) and MSS (microsatellite stable), which have ≥ 30 – 40 %, > 0 % – < 30 – 40 %, and 0 % unstable markers, respectively [Boland et al., 1998]. Immunohistochemical (IHC) analysis of MMR protein expression was performed on paraffin-embedded tumor tissue sections from 42 samples, as previously described [Roncari et al., 2007].

Allelic frequencies in control chromosomes

To assess frequencies of the 24 *MLH1* and *MSH2* variants in control chromosomes, one hundred and sixty DNA samples from anonymized healthy Italian blood donors with no history of colorectal cancer among 1st degree relatives and from the same region of origin (Tuscany) of most of the patients were analyzed by direct sequencing. In addition, the Exome Aggregation Consortium database (ExAC Browser (Beta), http://exac.broadinstitute.org/, 04/2016 accessed) was interrogated, excluding the Cancer Genome Atlas (TGCA) data.

Co-segregation with phenotype and multifactorial likelihood analysis

Co-segregation analysis was performed for 11 variants in 16 families; in these, 24 affected carrier relatives, in addition to probands, were identified. The variants were detected by direct sequencing. Multifactorial likelihood analysis was performed for 14 variants for which sufficient data were

available, as described previously [Thompson et al. 2013a; Thompson et al. 2013b]. Briefly, a probability of pathogenicity based on variant location or *in silico* scoring of missense substitutions is combined with likelihood ratios (LR) for segregation and tumour characteristics (MSI/BRAF status) to derive a posterior probability of pathogenicity.

mRNA splicing analysis

Total RNA was extracted from the 7 cycloheximide-treated and untreated LCLs established from *MLH1* or *MSH2* variant carriers, using RNeasy® Plus Mini Kit (Qiagen, Hilden, D)**.** Cycloheximide (Sigma-Aldrich, Saint Louis, MO, USA) was added at 10µg/ml to the medium 4 hr before harvesting the cells to prevent degradation of unstable transcripts by nonsense-mediated decay (NMD). cDNA was synthesized using TaqMan Reverse Trascription Kit (Applied Biosystems, Foster City, CA, USA). Primers and conditions used for cDNA amplification are available upon request. PCR products were analyzed on agarose gels, and individual bands, corresponding either to the full length or to the aberrantly spliced transcripts were excised and eluted using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) before amplification and direct sequencing. All RT-PCR experiments were performed in duplicate. Since alternative splicing is commonly observed for *MLH1* and *MSH2* [Genuardi et al. 1998], to improve the detection and interpretation of splicing aberrations, eight control samples were also analyzed [Thompson et al., 2015].

Allele-specific expression (ASE) analysis

Allele-specific expression (ASE) was investigated by Single Nucleotide Primer Extension (SNuPE) in 10 patients heterozygotes for the coding SNPs rs1799977 (*MLH1* c.655G>A) or rs4987188 (*MSH2* c.965G>A, as previously described [Crucianelli et al., 2014]. Total RNA was extracted from blood samples collected into PAXgene Blood RNA tubes (PreAnalytiX, Qiagen, Hilden, Germany), using

the PAXgene Blood RNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Samples from heterozygotes for the same SNPs who had no additional *MLH1* and *MSH2* sequence change were used as controls [Perera et al., 2010; Pastrello et al., 2011]. ASE was calculated after measuring peak heights in heterozygous samples [Castellsagué et al., 2010]. Values included in the 0.8-1.2 range were assumed as a cut-off for normal ASE according to previous studies [Renkonen et al., 2003; Castellsagué et al., 2010; Perera et al., 2010]. All experiments were carried out in triplicate and two independent replicates of all experiments were performed. Control heterozygotes for the *MLH1* or *MSH2* exonic polymorphisms (rs1799977 and rs4987188, respectively) were included in each experiment.

Minigene splicing assay

Splicing assays were performed by comparing the splicing pattern of WT and mutant minigenes transiently expressed in HeLa cells [Soukarieh et al., 2016]. Two different vectors were used in the minigene splicing assay: pCAS2 or pSPL3m [Soukarieh et al., 2016], as specified. Except for 2 contructs (*MSH2* c.2006-6T>C and c.2081T>C), minigenes were prepared by first PCR-amplifying wild-type (WT) and mutant genomic segments from patients' DNA using forward and reverse primers mapping approximately at 150 nucleotides upstream and downstream the exon of interest, respectively. Primer sequences are available upon request. The PCR products were then inserted into the intron of pCAS2 to generate splicing reporter minigenes as previously described [Tournier et al., 2008]. Minigenes carrying the single variants *MSH2* c.2006-6T>C and *MSH2* c.2081T>C (present in *cis* in patient genomic DNA) were prepared by site-directed mutagenesis by using a two-stage overlap extension PCR method [Ho et al., 1989]. The psPL3m construct carrying *MLH1* c.301G>A was generated by transferring the insert from the previously described minigene pCAS1.MLH1.exon3.c.301G>A [Tournier et al., 2008].

227

228 *Protein stability and vitro MMR activity analyses*

229 Altogether 9 *MLH1* and *MSH2* missense variations were introduced into the *MLH1* and *MSH2*
230 cDNAs cloned into a pFastBac1 vector (Invitrogen, Carlsbad, CA, USA), using a PCR-based site-
231 directed mutagenesis kit according to manufacturer´s instructions (QuikChange Lightning®Site-
232 directed mutagenesis Kit, Stratagene, La Jolla, CA, USA). The mutated constructs were sequenced
233 (ABIPrism 3100 Genetic Analyzer; Applied Biosystems, Foster City, CA, USA) prior to protein
234 production. Primer sequences and PCR parameters are available upon request.

235

236 Recombinant baculoviruses were generated by Bac-to-Bac system (Invitrogen, Carlsbad, CA, USA)
237 and proteins were produced in *Spodoptera frugiperda* (*Sf*9) insect cells as described earlier [Nyström-
238 Lahti et al., 2002; Kariola et al, 2002; Ollila et al., 2006]. For protein production *Sf9* cells were co-
239 infected with MLH1 and PMS2 viruses to yield MutLα heterodimers, or MSH2 and MSH6 viruses to
240 yield MutSα heterodimers. The total protein extracts (TE) were prepared as previously described
241 [Nyström-Lahti et al., 2002; Kariola et al., 2002].

242

243 The expression levels of produced protein variants were studied by Western blot analysis. The
244 proteins were blotted onto nitrocellulose membranes (Hybond, PVDF, Amersham Pharmacia biotech,
245 Uppsala, Sweden) and visualized with anti-MLH1 (BD Biosciences/Pharmingen, San Diego, CA,
246 USA, clone 168-15) (0.5 µg/ml), anti-PMS2 (Calbiochem/Oncogene Research, San Diego, CA, USA,
247 Ab-1) (0.2 µg/ml), anti-MSH2 (Calbiochem, San Diego, CA, USA, MSH2- Ab1, NA-26) (0.2 µg/ml)
248 and anti-MSH6 (BD Transduction Laboratories, Lexington, KY, USA, clone 44) (0.02 µg/ml)
249 monoclonal antibodies. To estimate the MMR protein level in the extracts, α-tubulin was used as a
250 loading control (anti-α-tubulin; Sigma, Louis, MO, USA, DM1A) (0.2 µg/ml).

The repair efficiencies (R%) of the recombinant protein variants were analyzed by complementing the MMR-deficient nuclear extracts (NE) of HCT116 or LoVo cells (American Type Culture Collection, Manassas, VA, USA) with the *Sf9* TEs containing overexpressed MLH1 or MSH2 proteins [Nyström-Lahti et al., 2002; Kantelinen et al., 2012] according to the protocol previously described [Kantelinen et al., 2012]. Seventy-five μg of NE was incubated with TE including in each sample comparable amounts of MutLα or MutSα, respectively, and with an excess amount (100 ng) of the heteroduplex DNA substrate. R% was calculated as an average of three independent experiments using GeneTools 3.08 (SynGene, Cambridge, England). The relative R% was calculated in respect to the WT control [Drost et al., 2010; Kantelinen et al., 2012].

Clinical classification of variants

The 5 class system for clinical classification recommended by the International Agency for Research on Cancer (IARC) working group on the interpretation of DNA sequence variants in cancer predisposition genes was used [Plon et al., 2008]. Class 5 and 4 include definitely pathogenic and likely pathogenic variants, respectively; when observed in a proband, they provide confirmation of LS diagnosis, so that relatives can be consequently offered predictive testing. Variants assigned to class 1 and 2 correspond to definitely and likely neutral/not pathogenic (or of low clinical impact) sequence changes, respectively; their detection is not followed by further clinical testing in the family. Finally, class 3 includes all those variants whose clinical and/or functional effects cannot be determined, due to either insufficient (e.g. detection in a single family) or contradictory evidence (ie discordant results

from similar *in vitro* assays performed in different laboratories); these are also called variants of uncertain (or unknown) significance (VUS).

RESULTS

Population frequency

We first verified variant allele frequencies in control populations (Table 1). Previous classifications performed by InSiGHT relied on frequencies reported in the 1000 Genomes Project database (http://www.1000genomes.org/), in addition to data published or reported by single centers on local populations. In this study we used values from the Exome Aggregation Consortium (ExAC), which is a more comprehensive dataset and which also incorporates the 1000 Genomes data. We also tested 160 Italian control subjects for 20 variants.

Four variants (*MLH1* c.1558+14G>A; *MSH2* c.380A>G, c.1511-9A>T and c.2006-6T>C) that reached minor allele frequencies $\geq$ 0.01 had previously been classified as Class 1-not pathogenic based on population frequency data alone [Thompson et al., 2014], and were subjected to further analyses to verify consistency across different points of evidence. The frequency of *MSH2* c.380A>G in phase 1 of the 1000 Genomes Project, which had been originally used for classification, was 0.02, while the currently reported frequency in ExAC is slightly lower, 0.00692. Of note, the frequencies of *MLH1* c.1039-8T>A were 0.00155 in ExAC and 0.02187 in 160 Italian controls, respectively. This difference is likely accounted for by a low quality of calling in the ExAC population, as suggested by the following observations: (i) it is called in less than 80% of individuals in ExAC; (ii) the variant was

297 found to be in linkage disequilibrium with *MLH1* c.1558+14G>A in the Italian population; and (iii)
298 *MLH1* c.1558+14G>A had similar polymorphic frequencies in ExAC and in Italian controls (0.03948
299 versus 0.02187).
300
301 *Tumor pathology data*
302 Analyses performed on tumors included microsatellite instability (MSI), MMR
303 immunohistochemistry (IHC), *BRAF* p.Val600Glu somatic mutation, *MLH1* promoter methylation,
304 and loss of heterozygosity (LOH) (Table 1).
305
306 MSI and/or IHC data were available for 23/24 and 22/24 variants, respectively (Table 1). These
307 included 12/13 predicted missense changes, for 6 of which $\geq$ 2 tumors were investigated. Results
308 indicative of *in vivo* MMR inactivation (MSI-H status and/or lack of expression of the protein
309 encoded by the gene carrying the constitutional variant) were observed for 7 predicted missense
310 substitution variants: *MLH1* c.301G>A p.(Gly101Ser), c.779T>G p.(Leu260Arg), c.1421G>C
311 p.(Arg474Pro), c.1814A>C p.(Glu605Ala), and *MSH2* c.2081T>C p.(Phe694Ser) and c.2087C>T
312 p.(Pro696Leu), as well as for the potential splice variant *MSH2* c.2006G>T. Results indicative of *in*
313 *vivo and in vitro* MMR proficiency (MSS status, normal MMR protein expression and proficient
314 functional assays) were obtained on 2 missense variants, *MLH1* c.1043T>C p.(Leu348Ser) and *MSH2*
315 c.244A>G p.(Lys82Glu). Discordant MSI and IHC results were observed for the missense variants
316 *MLH1* c.2041G>A p.(Ala681Thr) and c.2059C>T p.(Arg687Trp) in tumors from different
317 individuals; for both variants 1 MSI-H tumor showed normal IHC staining, while 1 MSS sample was
318 associated with lack of MLH1 expression.
319

BRAF and/or *MLH1* promoter analyses were performed on tumor samples for 9 variants (Table 1). The *BRAF* p.Val600Glu mutation was detected in two MLH1-negative tumors from carriers of the *MLH1* variants c.1421G>C p.(Arg474Pro) and c.1743G>A p.(Pro581=), respectively. *MLH1* epigenetic defects could be tested only for the tumor from the carrier of the c.1743G>A variant, which however did not show *MLH1* promoter hypermethylation. Four additional samples had both *BRAF* p.Val600Glu and *MLH1* promoter methylation tested: both analyses were negative in 3 MSI-H samples from carriers of *MLH1* c.779T>G p.(Leu260Arg) carriers and in 1 MSI-H sample from a c.2041G>A p.(Ala681Thr) carrier.

LOH analysis was performed for 14 variants, 8 *MLH1* and 6 *MSH2* (Table 1). Loss of the variant allele was detected in tumors from carriers of 3 different *MLH1* variants: c.1217G>A p.(Ser406Asn), c.1421G>C p.(Arg474Pro) and c.1732-19T>C.

RNA analyses

The *MSH2* variant c.2006G>T, which is currently assigned to Class 5 based on evidence of a major splicing defect, was associated with complete exclusion of exon 13 in the minigene assay (Fig. 1), but with only partial skipping in the LCL from a carrier (Fig. 2A and 2B). Both alleles at position 2006 were detected in the full-length cDNA product from the LCL (Fig. 2C and 2D). Partial exon 13 loss was also detected in blood samples drawn in PAXgene tubes from the patient above and from 2 additional individuals carrying *MSH2* c.2006G>T investigated in another laboratory in France, as well as in a LCL established from one of these French carriers (data not shown).

Three other variants (*MLH1* c.301G>A and c.1039-8T>A; *MSH2* c.2006-6T>C) were associated with partial exon skipping (Table 2; Fig. 1; Supp. Fig. S1-S2) in patient samples, in the *ex vivo* minigene assay, or in both. These involved in all cases known alternatively spliced transcripts [Genuardi et al., 1998; Thompson et al., 2015]. Interestingly, for *MSH2* c.2006-6T>C partial exon 13 exclusion was only detected by the minigene assay (Fig. 1) but not in the patient sample (Tournier et al., 2008), further suggesting that the minigene assay may overestimate the splicing defect for this exon. On the other hand, in this study partial skipping of exon 12 was observed in lymphoblastoid cells (data not shown) but not with the minigene assay for *MLH1* c.1039-8T>A (Supp Fig. S2). The latter result is in accordance with previous findings obtained for this variant by minigene assay [Petersen et al., 2013].

In addition, the minigene assay showed that Class 5-pathogenic *MLH1* c.301G>A p.(Gly101Ser) is associated with loss of the use of an alternative splice site (Supp Fig. S3); while the clinical significance of this finding cannot be established based on the minigene result only, the variant allele should produce only the canonical transcript.

None of the 10 variants tested by the SNuPE assay showed evidence of allelic expression imbalance, consistent with the splicing assay results (Table 2).

Mismatch repair activity and protein expression analyses

An *in vitro* MMR complementation assay based on the synthesis of MMR protein in *Spodoptera frugiperda* (*Sf*9) insect cells was performed in MMR-defective human cell lines for 9 coding variants (Table 2). Three of the five MMR defective protein variants (MLH1 p.Leu260Arg; MSH2 p.Phe694Ser and p.Pro696Leu were found to be unstable *in vitro* (Fig. 3). Loss of MMR activity

(relative activity < 1%) was demonstrated for 5 variants: MLH1 p.Gly101Ser and p.Leu260Arg, and MSH2 p.Gly669Val, p.Phe694Ser and p.Pro696Leu. The remaining four variants, MLH1 p.Leu348Ser, p.Arg474Pro, p.Glu605Ala, and MSH2 p.Lys82Glu, were all stable in the transient expression assay and MMR proficient.

Four other missense variants, MLH1 p.Ser406Asn, p.Ala681Thr, p.Arg687Trp, and MSH2 p.Asn127Ser, had been previously shown to be proficient in the MMR activity assay, although two of them, MLH1 p.Ala681Thr and p.Arg687Trp, showed discordant protein instability results across different studies (Table 2).

Clinical data, multifactorial analysis and variant classification

Family history types, co-occurrence of other MMR gene variants, and the components and results multifactorial analysis are shown in Supp. Tables S1-S2.

Multifactorial analysis was performed for 14 variants. Using quantitative analysis (based on multifactorial posterior probability) and/or assessment of qualitative criteria, variants were classified according to the 5-tier system proposed by InSiGHT (Table 3) [Thompson et al., 2014]. Four of the six novel variants (*MLH1*, c.1732-19T>C and c.1743G>A; *MSH2* c. 244A>G and c.2442T>G) were assigned to Class 2-likely not pathogenic. Of note, one novel variant, *MLH1* c.1814A>C p.(Glu605Ala), is in class 4-likely pathogenic according to multifactorial analysis based only on 2 available values, a 0.7 prior probability calculated *in silico*, and a single MSI-H CRC not expressing the MLH1 protein. The remaining novel variant, *MLH1* c.1043T>C p.(Leu348Ser) is in Class 3-VUS due to insufficient evidence.

The previous InSiGHT assignment of *MSH2* c.2006G>T to Class 5-pathogenic based on RNA splicing data was confirmed by the results of multifactorial analysis in this study (Table 3).

The classification made by InSiGHT was changed for 3 variants after the addition of novel segregation and molecular tumor data. *MLH1* c.301G>A p.(Gly101Ser), originally in Class 4-likely pathogenic, was upgraded to Class 5-pathogenic, while variants *MSH2* c.1387-8G>T and c.1737A>G p.(Lys579=) were moved from Class 2-likely not pathogenic to Class 1-not pathogenic.

Previous classifications of the other 15 variants were supported by the data obtained. Results of novel RNA analyses performed in this study were in agreement with the assignment to Class 1-not pathogenic of intronic *MLH1* variants c.1039-8T>A and c.1558+14G>A. Insufficient evidence to attain a clinically actionable category was available for *MLH1* c.1421G>C p.(Arg474Pro), even though novel data from tumor studies brought down the posterior probability of pathogenicity from 0.51 to 0.094.

DISCUSSION

The ultimate purpose of genetic testing for LS and other cancer predisposition syndromes is to reduce cancer morbidity and mortality through the implementation of specific preventive options for carriers of disease causing variants. Interpreting the significance of DNA variants identified in the diagnostic

laboratory is an integral component of clinical DNA testing. The interpretation process is complex, as several independent datasets must be taken into account. Recently, recommendations for clinical classification of MMR gene variants have been formulated [Thompson et al., 2014]. We have performed a thorough investigation of 24 MMR gene sequence variants identified in a single center in order to assess their clinical relevance, using points of evidence that are included in the InSiGHT recommendations, as well as additional potential classification components. Our findings confirm the overall validity of the InSiGHT criteria and suggest that the interpretation process could be improved by introducing some adjustments.

Overall, our results provide support to or improve previous classifications for the 18 variants that had already been assessed by InSiGHT (http://insight-group.org/variants/classifications). For 3 of these variants (*MLH1* c.301G>A; *MSH2* c.1387-8G>T and c.1737A>G), a more stable classification, either from Class 4-likely pathogenic to Class 5-pathogenic or from Class 2-likely not pathogenic to Class 1-not pathogenic, was achieved using novel clinical and molecular data. These changes do not affect cancer prevention strategies, since the same clinical recommendations apply to Class 5 and 4 and Class 2 and 1, respectively [Plon et al., 2008]. However, assignments to Class 5 and 1 can be considered definitive, since the likelihood that a variant in either of these categories will be moved to a class associated with different clinical advice is very low [Plon et al., 2008]. The IARC/InSiGHT interpretation criteria advise to consider research testing of further samples/relatives to try and obtain definitive classifications for Class 4-likely pathogenic and Class 2-likely not pathogenic variants [Plon et al. 2008; Thompson et al. 2014], and our results demonstrate the practical importance of this recommendation. For all 3 variants the classification was based on multifactorial analysis, and in all cases the class switch was made possible by the incorporation of few novel data on tumor characteristics and/or segregation, highlighting the relevance of collecting these types of information.

The novel variants *MLH1* c.1814A>C p.(Glu605Ala) and *MSH2* c.244A>G p.(Lys82Glu) were in Class 4-likely pathogenic and Class 2-likely not pathogenic, respectively, following multifactorial analysis. For both variants only one clinical observation, that is, molecular information obtained on a single tumor sample (Table 1 and Supp. Table S2), is available. The Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) in the *BRCA1/BRCA2* (MIM# 113705 and MIM# 600185) genes recommends that variants attaining thresholds for assignment to clinically actionable classes by multifactorial analysis with limited contribution from clinical or laboratory evidence be considered of uncertain significance until further evidence is accrued (http://enigma.consortium.org/documents/ENIGMA_Rules_2015_03_26.pdf). We propose to adopt this recommendation also for the MMR genes, especially when there is apparent discordance between functional and clinical evidence, such as for *MLH1* p.(Glu605Ala). In particular, for *MLH1* variants additional evidence from *BRAF* and/or promoter methylation tumor testing could be used to reinforce the evidence in favor of pathogenicity.

The partially discordant RNA splicing results between the minigene assay and analyses of patient derived samples obtained in this study for *MSH2* c.2006G>T suggest that the splicing alteration may not be the only or the major inactivation mechanism for *MSH2* c.2006T. Indeed, the functional *in vitro* assay showed reduced repair activity of the protein encoded by the variant allele, p.669Val and the variant could be assigned to Class 5 also based on multifactorial analysis. However, complete absence of c.2006T allele in full-length transcript.in patient RNA, together with total exon 13 exclusion in a minigene assay, was observed in another study [van der Klift et al., 2015]. Therefore, further studies will be needed to clarify the mechanisms underlying pathogenicity of *MSH2* c.2006G>T. At the same time, the interpretation criteria for RNA analyses should be reconsidered based on these apparently inconsistent results.

Minor effects on splicing were observed either on patient RNA or by the minigene assay in this study for the Class 1-not pathogenic variants *MLH1* c.1039-8T>A and *MSH2* c.380A>G and c.2006-6T>C. All are in Class 1 based on population frequency only, confirming that they have no major clinical effects [Genuardi et al., 1998; Thompson et al., 2015].

None of the other variants were found to be associated with significant splicing anomalies. Lack of abnormal splicing products was important to assign *MLH1* c.307-19A>G, c.1732-19T>C and c.1743G>A, and *MSH2* c. 2442T>G p.(Leu814=) to Class 2-likely not pathogenic. Two of them, *MLH1* c.307-19A>G and c.1743G>A also had population frequencies ~1/5,000 and ~1/10,000, respectively. According to the InSiGHT criteria, synonymous or deep intronic variants for which splicing assays do not show alterations should be considered as Class 2-likely not pathogenic. One of the combinations required for assignment of a variant to Class 1-not pathogenic includes all of the following points of evidence: allelic frequency 0.01%-1%, lack of co-segregation with disease, estimated risk <1.5 determined by case-control studies, and presence of molecular features not compatible with involvement of the gene carrying the variant in ≥ 3 tumors; this criterion applies to all types of variants, regardless of their nature and prior probability of altering gene function. Since intronic and synonymous variants have a low *a priori* likelihood of affecting gene processing, combinations of any of the above evidences (ie, lack of segregation, population frequency, low estimated risk, and molecular characteristics) and normal splicing patterns could reasonably be considered sufficient for assignment to Class 1. Interestingly, the association of intronic location or synonymous coding nucleotide substitution and absence of mRNA aberrations demonstrated by *in vitro* assays has been proposed by the ENIGMA consortium as a criterion for assignment of variants in the *BRCA1/BRCA2* genes to Class 1-not pathogenic (http://enigmaconsortium.org/documents/ENIGMA_Rules_2015-03-26.pdf). Data from our study

indicate that it would be justified to consider homogenization of the Class 1 criteria for intronic/synonymous substitutions between the ENIGMA and InSiGHT consortia.

Nine missense variants were investigated by *in vitro* MMR assay based on the production of MSH2 or MLH1 proteins in *Sf9* insect cells and subsequent complementation of human MMR deficient cell lines. The same assay had been previously used for three other missense substitutions found in our series [Raevaara et al., 2005; Ollila et al., 2008; Christensen et al., 2009; Kansikas et al., 2011], while one variant - *MLH1* c.1217G>A p.(Ser406Asn) - had been tested with two different mammalian repair assays. All five MMR deficient variants (*MLH1* p.(Gly101Ser) and p.(Leu260Arg); *MSH2* p.(Gly669Val), p.(Phe694Ser) and p.(Pro696Leu)) are in Class 5-pathogenic, supporting the classifications based on multifactorial analysis. Of note, the aminoacids replaced in *MSH2* p.(Phe694Ser) and p.(Pro696Leu) are located nearby in the ATPase domain, indicating that this region is particularly sensitive to structural changes; this suggestion is reinforced by the observation that none of the 29 reported *MSH2* exon 13 missense changes have been so far assigned to Class 1-not pathogenic or Class 2-likely not pathogenic by InSiGHT (http://chromium.lovd.nl/LOVD2/colon_cancer/variants.php?select_db=MSH2&action=search_all&search_Variant%2FExon=13&search_MutCol=%3E&search_Variant%2FDNA=&search_Variant%2FRNA=&search_Variant%2FProtein=&search_Patient%2FPhenotype%2FDisease=&search_Patient%2FReference=).

MLH1 c.2041G>A p.(Ala681Thr) and c.2059C>T p.(Arg687Trp) are assigned to Class 5-pathogenic despite the results of the functional assays, which show inconclusive data on protein expression and normal MMR activity, with discordant observations across different studies. However, both are associated with an abundance of clinical data allowing them to overcome the Class 5-pathogenic

posterior probability threshold using multifactorial analysis. At the same time discordant tumor pathology findings, including samples that were MSS and/or expressed MLH1, have also been reported for both variants. It will be interesting to verify the degree of phenotypic expression associated with these two variants. By analogy with equivocal functional results obtained on the BRCA1 variant p.Arg1699Gln [Spurdle et al., 2012] they might be considered as candidate intermediate risk variants. Notably, other *MLH1* missense substitutions located in proximity of these variants are associated with proficient repair but reduced or inconclusive protein expression data; these include for instance the Class 5-pathogenic c.1942C>T p.(Pro648Ser) and c.1943C>T p.(Pro648Leu), and the Class 3-VUS c.1918C>T p.(Pro640Ser), c.1919C>T p.(Pro640Leu), c.1976G>A p.(Arg659Gln), c.2027T>G p.(Leu676Arg), and c.2027T>C p.(Leu676Pro). Therefore variants located in this region of the MLH1 protein may cause functional impairment through reduced expression/stability and/or other as yet to be determined mechanisms not directly affecting repair activity.

While LOH is an important silencing mechanism of the wild type allele [Alemayehu et al., 2007], so far it has not been considered as a point of evidence for MMR gene variant classification by InSiGHT. This is due to several reasons, including multiple observations of loss of variant pathogenic alleles in cancers from MMR gene carriers and technical difficulties, ie, due to the potential presence of MSI hampering analysis of LOH using microsatellite markers [Hofstra et al., 2008]. The findings from this study, especially the observation of loss of the variant allele in samples from carriers of Class 1-not pathogenic and Class 2–likely not pathogenic variants confirm that LOH should be considered with caution for the interpretation of variant pathogenicity in the MMR genes. Studies on large series are needed to assess the usefulness of this marker and its predictive value.

The evaluation of multiple clinical parameters and functional assays undertaken in this study allows refining the strategy for the clinical classification of MMR gene variants. Intronic and synonymous variants that cannot be tested in the *in vitro* MMR assay should be assessed for effects on RNA processing, by detection of aberrant transcripts (in the presence of NMD inhibitors) and allele-specific expression (in the absence of NMD inhibitors). We suggest that, when no major alteration is observed, the variant could be assigned to Class 1-not pathogenic, even without further evidence (from ie, segregation and tumor characteristics), as stated by ENIGMA for *BRCA1* and *BRCA2*. The underlying rationale is that the probability that an intronic variant with no documented splicing aberration will cause high tumor risk is very low, < 1/1,000.

For potential missense variants, concordant evidence in favour or against pathogenicity should be derived both from functional assays - RNA first, and if normal, protein - and clinical data. Given the importance of obtaining segregation and molecular tumor results for the purpose of variant classification, any attempt should be made to test additional patients and samples, especially from carriers of missense variants which are usually more difficult to classify compared to silent and intronic changes.

Finally, classifications obtained by multifactorial analysis should be supported by multiple data points; this could be achieved by requiring a minimum threshold or different points of evidence from clinical and tumor data to allow assignment to a clinically actionable class.

humu23117-sup-0001-SuppMat.docx

Supplementary Information Figure S1. Identification of *MSH2* splicing effects by using a splicing minigene reporter assay. (A) Structure of the pCAS2-MSH2 minigenes. Boxes represent

exons and lines in between indicate introns. *MSH2* segments are shown in dark colour. Splicing events detected in the minigene assay are indicated by the dotted lines and further described on the right. (B) Analysis of the splicing pattern of wild-type and mutant pCAS2-MSH2 minigene transcripts. Wild-type and mutant minigenes, as indicated, were transfected into HeLa cells and the minigene transcripts were analyzed by RT-PCR. The image shows a 2% agarose gel stained with ethidium bromide, visualized by exposure to ultraviolet light under conditions of non-saturating exposure. The identities of the RT-PCR products are shown both on the left and on the right of the gel. Results are representative of 3 independent experiments. M, size marker (100 bp DNA ladder, New England Biolabs); pCAS2, empty vector; WT, wild-type.

Supplementary Information Figure S2. Analysis of the impact on splicing of *MLH1* variants by using a minigene reporter assay. (A) Structure of the pCAS2-MLH1 minigenes. Boxes represent exons and horizontal lines in between indicate introns. *MLH1* segments are shown in dark colour. Splicing events detected in the minigene assay are indicated by the dotted lines and further described on the right. (B) Analysis of the splicing pattern of wild-type and mutant pCAS2-MLH1 minigene transcripts. Wild-type and mutant minigenes, as indicated, were transfected into HeLa cells and the minigene transcripts were analyzed by RT-PCR. The image shows a 2% agarose gel stained with ethidium bromide, visualized by exposure to ultraviolet light under conditions of non-saturating exposure. Results are representative of 3 independent experiments. M, size marker (100 bp DNA ladder, New England Biolabs); pCAS2, empty vector; WT, wild-type.

Supplementary Information Fig. S3. *MLH1* c.301G>A alters the alternative splicing pattern of *MLH1* exon 3 in the minigene splicing assay. (A) Structure of the pSPL3m-MLH1ex3 minigene. Boxes represent exons and horizontal lines in between indicate introns. The *MLH1* segment is shown in dark colour. Splicing events detected in the minigene assay are indicated by the dotted lines. (B) Analysis of the splicing pattern of wild-type and mutant pSPL3m-MLH1ex3 minigene transcripts. Wild-type and mutant minigenes, as indicated, were transfected into HeLa cells and the minigene transcripts were analyzed by RT-PCR as described under Materials and Methods. The image shows a 2% agarose gel stained with ethidium bromide, visualized by exposure to ultraviolet light under conditions of non-saturating exposure. The identities of the RT-PCR products are shown on the left and below the gel. (C) Usage of the reference 5'ss and a5'ss of *MLH1* exon 3 (NM_000249.3 and NM_001167617.1, respectively) in the WT and mutant contexts. The upper panel shows *in silico* predictions for the effect of c.301G>A on the strength of the 5' splice site of *MLH1*exon 3 (predictions obtained with 5 different algorithms, as described in Soukarieh et al., 2016). The bottom panel shows the sequence of the RT-PCR products indicated by the star (heteroduplexes) and purified from the gel shown in B. 5'ss, 5'splice site; a5'ss, alternative 5' splice site; Δ5 nts, deletion of the last 5 nucleotides of *MLH1* exon 3.

Supplementary Information Table S1. Clinical data and co-occurrence of multiple variants in families with MMR gene variants.

Supplementary Information Table S2. Segregation and multifactorial likelihood analysis for the investigated MMR gene variants.

ACKNOWLEDGMENTS

MG has been supported by a grant from Istituto Toscano Tumori (ITT). BAT is supported by an NHMRC Early Career Fellowship (ID1091211). ABS is supported by an NHMRC Senior Research Fellowship (ID1061779). Aspects of this research (bioinformatic interpretation) were supported by an NIH subcontract (grant ID NIH R01CA164944). MN has been supported by a grant from the European Research Council (2008-AdG-232635). Part of this project was supported by a grant from the French Institut National du Cancer/Direction Générale de l'Offre de Soins (INCa/DGOS) and the Fondation ARC pour la Recherche sur le Cancer to AM. OS was funded by a fellowship from the French Ministry of Education. The authors declare they have no conflict of interest.

InSiGHT VIC CONTRIBUTORS. *Kiwamu Akagi*, Div. Molecular Diagnosis & Cancer Prevention Saitama Cancer Center, Saitama, Japan. *Fahd Al-Mullah*, Molecular Pathology Unit, Health Sciences Center, Faculty of Medicine, Kuwait University, Safat Kuwait, Kuwait. *Ian R Berry*, Leeds Genetics Laboratory, Leeds, UK. *Michael Farrell*, Department of Cancer Genetics, Mater Private Hospital, Dublin, Ireland. *Susan Farrington*, Institute of Genetics and Molecular Medicine, University of Edinburgh, UK. *Ian Frayling*, Institute of Cancer & Genetics, Cardiff University, Cardiff, UK. *Elke Holinski-Feder*, Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München; and MGZ – Medizinisch Genetisches Zentrum, Munich, Germany. *Maija Kohonen-Corish*, The Kinghorn Cancer Centre, Garvan Institute of Medical Research, Sydney; St Vincent's Clinical School, UNSW Australia, Sydney; and School of Medicine, Western Sydney University, Sydney, Australia. *Kristina Lagerstedt-Robinson*, Department of Molecular Medicine and Surgery, Karolinska Institutet, and Department of Clinical Genetics, Karolinska University Hospital, Solna, Stockholm, Sweden. *Finlay Macrae*, Dept of Colorectal Medicine and Genetics, The Royal Melbourne Hospital, Victoria, Australia. *Pål Møller*, Research Group Inherited Cancer, Department of Medical Genetics, The Norwegian Radium Hospital, Oslo University Hospital, Norway. *Monika Morak*, Medizinische

Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München; and MGZ – Medizinisch Genetisches Zentrum, Munich, Germany. *John-Paul Plazzer*, Department of Colorectal Medicine and Genetics, Royal Melbourne Hospital, Melbourne, Victoria, Australia. *Lene Rasmussen*, Faculty of Health Sciences, University of Copenhagen. *Brigitte Royer-Pokora*, Institut für Humangenetik und Anthropologie Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany. *Rodney J. Scott*, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle; Information Based Medicine Program, Hunter Medical Research Institute, Newcastle; Division of Molecular Medicine, Pathology North (Newcastle), John Hunter Hospital, Australia. *Rolf H. Sijmons*, Dept of Genetics, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

REFERENCES

Alemayehu A, Tomkova K, Zavodna K, Ventusova K, Krivulcik T, Bujalkova M, Bartosova Z, Fridrichova I. 2007. The role of clinical criteria, genetic and epigenetic alterations in Lynch-syndrome diagnosis. Neoplasma 54:391-401.

Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, Tavtigian SV, Goldgar DE, Young JP, Spurdle AB. 2009. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 30:757-770.

Auclair J, Busine MP, Navarro C, Ruano E, Montmain G, Desseigne F, Saurin JC, Lasset C, Bonadona V, Giraud S, Puisieux A, Wang Q. 2006. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. Hum Mutat 27:145-154.

Betz B, Theiss S, Aktas M, Konermann C, Goecke TO, Möslein G, Schaal H, Royer-Pokora B. 2010. Comparative in silico analyses and experimental validation of novel splice site and missense mutations in the genes MLH1 and MSH2. J Cancer Res Clin Oncol 136:123-134.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. 1998. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 58:5248-5257.

Buhard O, Cattaneo F, Wong YF, Yim SF, Friedman E, Flejou JF, Duval A, Hamelin R. 2006. Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors. J Clin Oncol 24:241-251.

665 Castellsagué E, González S, Guinó E, Stevens KN, Borràs E, Raymond VM, Lázaro C, Blanco I,
666 Gruber SB, Capellá G. 2010. Allele-specific expression of APC in adenomatous polyposis families.
667 Gastroenterology 139:439-447.

668

669 Christensen LL, Kariola R, Korhonen MK, Wikman FP, Sunde L, Gerdes AM, Okkels H,
670 Brandt CA, Bernstein I, Hansen TV, Hagemann-Madsen R, Andersen CL, Nyström M,
671 Ørntoft TF. 2009. Functional characterization of rare missense mutations in MLH1 and
672 MSH2 identified in Danish colorectal cancer patients. Fam Cancer 8:489-500.

673

674 Couch FJ, Rasmussen LJ, Hofstra R, Monteiro AN, Greenblatt MS, de Wind N. IARC
675 Unclassified Genetic Variants Working Group. 2008. Assessment of functional effects of
676 unclassified genetic variants. Hum Mutat 29:1314-1326.

677

678 Crucianelli F, Tricarico R, Turchetti D, Gorelli G, Gensini F, Sestini R, Giunti L, Pedroni M, Ponz de
679 Leon M, Civitelli S, Genuardi M. 2014. MLH1 constitutional and somatic methylation in patients
680 with MLH1 negative tumors fulfilling the revised Bethesda criteria. Epigenetics 9(10):1431-1438.

681

682 den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, Roux AF,
683 Smith T, Antonarakis SE, Taschner PE. 2016. HGVS Recommendations for the description of
684 sequence variants: 2016 update. Hum Mutat. 37:564-569.

685

686 Di Giacomo D, Gaildrat P, Abuli A, Abdat J, Frébourg T, Tosi M, Martins A. 2013. Functional
687 analysis of a large set of BRCA2 exon 7 variants highlights the predictive value of hexamer scores in
688 detecting alterations of exonic splicing regulatory elements. Hum Mutat 34:1547-1557.
689
690 Drost M1, Zonneveld Je, van Dijk L, Morreau H, Tops CM, Vasen HF, Wijnen JT, de Wind
691 N. 2010. A cell-free assay for the functional analysis of variants of the mismatch repair
692 protein MLH1. Hum Mutat 31(3):247-253.
693
694 Furukawa T, Konishi F, Masubuchi S, Shitoh K, Nagai H, Tsukamoto T. 2002. Densely
695 methylated MLH1 promoter correlates with decreased mRNA expression in sporadic
696 colorectal cancers. Genes Chromosomes Cancer 35:1-10.
697
698 Gaildrat P, Killian A, Martins A, Tournier I, Frébourg T, Tosi M. 2010. Use of splicing reporter
699 minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol Biol
700 653:249-257.
701
702 Genuardi M, Viel A, Bonora D, Capozzi E, Bellacosa A, Leonardi F, Valle R, Ventura A, Pedroni M,
703 Boiocchi M, Neri G. 1998. Characterization of MLH1 and MSH2 alternative splicing and its
704 relevance to molecular testing of colorectal cancer susceptibility. Hum Genet 102:15-20.
705

706 Giunti L, Cetica V, Ricci U, Giglio S, Sardi I, Paglierani M, Andreucci E, Sanzo M, Forni M,
707 Buccoliero AM, Genitori L, Genuardi M. 2009. Type A micro satellite instability in pediatric gliomas
708 as an indicator of Turcot syndrome. Eur J Hum Genet 17:919-927.
709
710 Gylling A, Abdel-Rahman WM, Juhola M, Nuorva K, Hautala E, Jarvinen HJ, Mecklin JP, Aarnio M,
711 Peltomaki P. 2007. Is gastric cancer part of the tumour spectrum of hereditary non-polyposis
712 colorectal cancer? A molecular genetic study. Gut 56: 926–33.
713
714 Hardt K, Heick SB, Betz B, Goecke T, Yazdanparast H, Küppers R, Servan K, Steinke V,
715 Rahner N, Morak M, Holinski-Feder E, Engel C et al. 2011. Missense variants in hMLH1
716 identified in patients from the German HNPCC consortium and functional studies. Fam
717 Cancer 10: 273-284.
718
719 Hinrichsen I, Brieger A, Trojan J, Zeuzem S, Nilbert M, Plotz G. 2013. Expression defect
720 size among unclassified MLH1 variants determines pathogenicity in Lynch syndrome
721 diagnosis. Clin Cancer Res 19: 2432-2441.
722
723 Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis
724 byoverlap extension using the polymerase chain reaction. Gene 77:51–59.
725

726 Hofstra RM, Spurdle AB, Eccles D, Foulkes WD, de Wind N, Hoogerbrugge N, Hogervorst FB,
727 IARC Unclassified Genetic Variants Working Group. 2008. Tumor characteristics as an analytic tool
728 for classifying genetic variants of uncertain clinical significance. Hum Mutat 29:1292-1303.
729
730 Jakubowska A, Górski B, Kurzawski G, Debniak T, Hadaczek P, Cybulski C, Kladny J,
731 Oszurek O, Scott RJ, Lubinski J. 2001. Optimization of experimental conditions for RNA-
732 based sequencing of MLH1 and MSH2 genes. Hum Mutat 17:52-60.
733
734 Janssen MJ, Waanders E, Te Morsche RH, Xing R, Dijkman HB, Woudenberg J, Drenth JP. 2011.
735 Secondary, somatic mutations might promote cyst formation in patients with autosomal dominant
736 polycystic liver disease. Gastroenterology 141:2056-2063.
737
738 Janssen MJ, Salomon J, Te Morsche RH, Drenth JP. 2012. Loss of heterozygosity is present in SEC63
739 germline carriers with polycystic liver disease. PLoS One 7(11):e50324.
740
741 Kansikas M, Kariola R, Nyström M. 2011. Verification of the three-step model in assessing
742 the pathogenicity of mismatch repair gene variants. Hum Mutat 32:107-115.
743
744 Kantelinen J, Kansikas M, Candelin S, Hampel H, Smith B, Holm L, Kariola R, Nyström M.
745 2012. Mismatch repair analysis of inherited MSH2 and/or MSH6 variation pairs found in
746 cancer patients. Hum Mutat 33:1294-1301.

747

748 Kariola R, Raevaara TE, Lönnqvist KE, Nyström-Lahti M. 2002. Functional analysis of
749 MSH6 mutations linked to kindreds with putative hereditary non-polyposis colorectal cancer
750 syndrome. Hum Mol Genet 11:1303-1310.

751

752 Lucci-Cordisco E, Zito I, Gensini F, Genuardi M. 2003. Hereditary nonpolyposis colorectal
753 cancer and related conditions. Am J Med Genet A 122A:325-334.

754

755 Lynch HT, de la Chapelle A. 2003. Hereditary colorectal cancer. N Engl J Med 348:919-932.

756

757 Mancini I, Santucci C, Sestini R, Simi L, Pratesi N, Cianchi F, Valanzano R, Pinzani P, Orlando C.
758 2010. The use of COLD-PCR and high-resolution melting analysis improves the limit of detection of
759 KRAS and BRAF mutations in colorectal cancer. J Mol Diagn 12:705-711.

760

761 Montera M, Resta N, Simone C, Guanti G, Marchese C, Civitelli S, Mancini A, Pozzi S, De
762 Salvo L, Bruzzone D, Donadini A, Romio L, et al. 2000. Mutational germline analysis of
763 hMSH2 and hMLH1 genes in early onset colorectal cancer patients. J Med Genet 37(7):E7.

764

765 Niroula A, Vihinen M. Classification of Aminoacid Substitutions in Mismatch Repair Proteins Using
766 PON-MMR2. 2015. Hum Mutat 36:1128–1134.

767

768 Nyström-Lahti M, Perrera C, Räschle M, Panyushkina-Seiler E, Marra G, Curci A,
769 Quaresima B, Costanzo F, D'Urso M, Venuta S, Jiricny J. 2002. Functional analysis of
770 MLH1 mutations linked to hereditary nonpolyposis colon cancer. Genes Chromosomes
771 Cancer 33:160-167.

772

773 Ollila S, Sarantaus L, Kariola R, Chan P, Hampel H, Holinski-Feder E, Macrae F, Kohonen-
774 Corish M, Gerdes AM, Peltomäki P, Mangold E, de la Chapelle A, Greenblatt M, Nyström
775 M. 2006. Pathogenicity of MSH2 missense mutations is typically associated with impaired
776 repair capability of the mutated protein. Gastroenterology 131:1408-1417.

777

778 Ollila S, Dermadi Bebek D, Greenblatt M, Nyström M. 2008. Uncertain pathogenicity of
779 MSH2 variants N127S and G322D challenges their classification. Int J Cancer 123(3):720-
780 724.

781

782 Pastrello C, Pin E, Marroni F, Bedin C, Fornasarig M, Tibiletti MG, Oliani C, Ponz de Leon M, Urso
783 ED, Della Puppa L, Agostini M, Viel A. 2011. Integrated analysis of unclassified variants in
784 mismatch repair genes. Genet Med 13:115-124.

785

786 Perera S, Li B, Tsitsikotas S, Ramyar L, Pollett A, Semotiuk K, Bapat B. 2010. A novel and rapid
787 method of determining the effect of unclassified MLH1 genetic variants on differential allelic
788 expression. J Mol Diagn 12:757-764.
789
790 Petersen SM, Dandanell M, Rasmussen LJ, Gerdes AM, Krogh LN, Bernstein I, Okkels H, Wikman
791 F, Nielsen FC, Hansen TV. 2013. Functional examination of MLH1, MSH2, and MSH6 intronic
792 mutations identified in Danish colorectal cancer patients. BMC Med Genet 14:103.
793
794 Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FB,
795 Hoogerbrugge N, Spurdle AB, Tavtigian SV; IARC Unclassified Genetic Variants Working Group.
796 2008. Sequence variant classification and reporting: recommendations for improving the
797 interpretation of cancer susceptibility genetic test results. Hum Mutat 29:1282-1291.
798
799 Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lönnqvist KE, Holinski-Feder E,
800 Sutter C, McKinnon W, Duraisamy S, Gerdes AM, Peltomäki P, Kohonen-Ccorish M,
801 Mangold E, Macrae F, Greenblatt M, de la Chapelle A, Nyström M. 2005. Functional
802 significance and clinical phenotype of nontruncating mismatch repair variants of MLH1.
803 Gastroenterology 129:537-549.
804
805 Renkonen E, Zhang Y, Lohi H, Salovaara R, Abdel-Rahman WM, Nilbert M, Aittomaki K, Jarvinen
806 HJ, Mecklin JP, Lindblom A, Peltomaki P. 2003. Altered expression of MLH1, MSH2, and MSH6 in
807 predisposition to hereditary nonpolyposis colorectal cancer. J Clin Oncol 21:3629-3637.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405-424.

Roncari B, Pedroni M, Maffei S, Di Gregorio C, Ponti G, Scarselli A, Losi L, Benatti P, Roncucci L, De Gaetani C, Camellini L, Lucci-Cordisco E, Tricarico R, Genuardi M, Ponz de Leon M. 2007. Frequency of constitutional MSH6 mutations in a consecutive series of families with clinical suspicion of HNPCC. Clin Genet 72:230-237.

Sijmons RH, Greenblatt MS, Genuardi M. Gene variants of unknown clinical significance in Lynch syndrome. An introduction for clinicians. 2013. Fam Cancer 12:181-187.

Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, Tosi M, Martins A. 2016. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools. PLoS Genet 12:e1005756.

827 Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM; IARC Unclassified
828 Genetic Variants Working Group. 2008. Prediction and assessment of splicing alterations:
829 implications for clinical testing. Hum Mutat 29:1304-1313.

830

831 Suraweera N, Duval A, Reperant M, Vaury C, Furlan D, Leroy K, Seruca R, Iacopetta B, Hamelin R.
832 2002. Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide
833 repeats and pentaplex PCR. Gastroenterology 123:1804-1811.

834

835 Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. 2007.
836 Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays.
837 Cancer Res 67: 4595-4604.

838

839 Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. 2008. Classification of rare missense
840 substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications.
841 Hum Mutat 29:1342-54.

842

843 Tavtigian SV, Greenblatt MS, Goldgar DE, Boffetta P; IARC Unclassified Genetic Variants
844 Working Group. 2008. Assessing pathogenicity: overview of results from the IARC
845 Unclassified Genetic Variants Working Group. Hum Mutat 29:1261-1264.

846

847 Thompson D, Easton DF, Goldgar DE. 2003. A full-likelihood method for the evaluation of causality
848 of sequence variants from family data. Am J Hum Genet 73:652-655.
849
850 Thompson BA, Goldgar DE, Paterson C, Clendenning M, Walters R, Arnold S, Parsons MT, Michael
851 D W, Gallinger S, Haile RW, Hopper JL, Jenkins MA et al. 2013a. A multifactorial likelihood model
852 for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and
853 tumor characteristics: a report from the Colon Cancer Family Registry. Hum Mutat 34:200-209.
854
855 Thompson BA, Greenblatt MS, Vallee MP, Herkert JC, Tessereau C, Young EL, Adzhubey IA, Li B,
856 Bell R, Feng B, Mooney SD, Radivojac P et al. 2013b. Calibration of multiple in silico tools for
857 predicting pathogenicity of mismatch repair gene missense substitutions. Hum Mutat 34:255-265.
858
859 Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, Bapat B, Bernstein I,
860 Capellá G, den Dunnen JT, du Sart D, Fabre A et al. 2014. Application of a 5-tiered scheme for
861 standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-
862 specific database. Nat Genet 46:107-115.
863
864 Thompson BA, Martins A, Spurdle AB. 2015. A review of mismatch repair gene transcripts: issues
865 for interpretation of mRNA splicing assays. Clin Genet 87:100-108.
866
867 Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, Wang Q, Buisine
868 MP, Soret J, Tazi J, Frébourg T, Tosi M. 2008. A large fraction of unclassified variants of the

mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Hum Mutat 29:1412-1424.

van der Klift H, Jansen AML, van der Steenstraten N, Bik EC, Tops CMJ, Devilee P, Wijnen JT. 2015. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. Mol Genet Genom Med 3:327-345.

Vasen HF. 2005. Clinical description of the Lynch syndrome [hereditary nonpolyposis colorectal cancer (HNPCC)]. Fam Cancer 4:219-225.

Vasen HF, Blanco I, Aktan-Collan K, Gopie JP, Alonso A, Aretz S, Bernstein I, Bertario L, Burn J, Capella G, Colas C, Engel C et al. 2013. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut 62:812-823.

Xie J, Guillemette S, Peng M, Gilbert C, Buermeyer A, Cantor SB. 2010. An MLH1 mutation links BACH1/FANCJ to colon cancer, signaling, and insight toward directed therapy. Cancer Prev Res (Phila) 3:1409-1416.

FIGURE LEGENDS

Figure 1. Identification of *MSH2* exon 13 splicing alterations by using a splicing minigene reporter assay. (A) Structure of the pCAS2-MSH2ex13 minigene. Boxes represent exons and horizontal lines in between indicate introns. The *MSH2* segment is shown in dark colour. Splicing events detected in the minigene assay are indicated by the dotted lines and further described on the right. (B) Analysis of the splicing pattern of wild-type and mutant pCAS2-MSH2ex13 minigene transcripts. Wild-type and mutant minigenes, as indicated, were transfected into HeLa cells and the minigene transcripts were analyzed by RT-PCR as described under Materials and Methods. The image shows a 2% agarose gel stained with ethidium bromide, visualized by exposure to ultraviolet light under conditions of non-saturating exposure. The identities of the RT-PCR products are shown on the right. Results are representative of 3 independent experiments. Marker, 100 bp DNA ladder (New England Biolabs); pCAS2, empty vector; WT, wild-type; *, heteroduplexes.

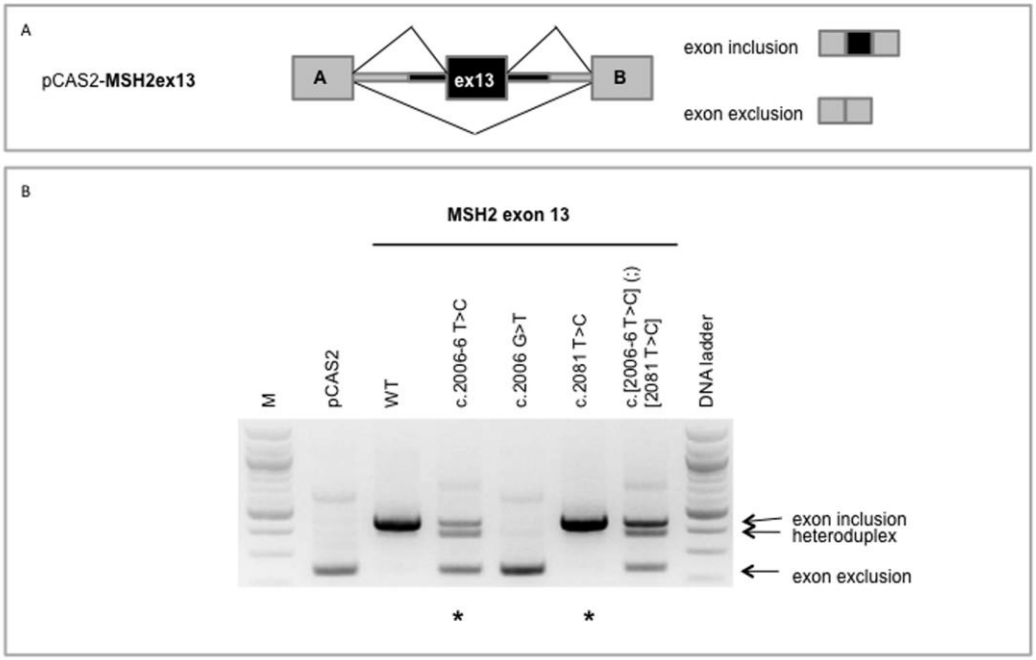


Figure 2. Splicing analysis by RT-PCR on cDNA from a LCL established from a carrier of *MSH2* variant c.2006G>T. (A) Gel electrophoresis of cDNA PCR products obtained using primers located in *MSH2* exons 12 and 14. MW: molecular weight marker (100 bp ladder). The upper band corresponds to the full length mRNA product, the fainter lower band to the isoform lacking exon 13. The arrow next to the asterisk shows the heteroduplex formed by the two PCR products. (B) Schematic representation of *MSH2* mRNA encompassing exons 12-14 and of the two splicing products detected in the LCL sample. (C) Electropherogram of the sequence of the Δ13 cDNA product. (D) Electropherogram of the sequence of the full-length cDNA product showing presnece of both alleles at c.2006G>T (the reverse strand is shown).

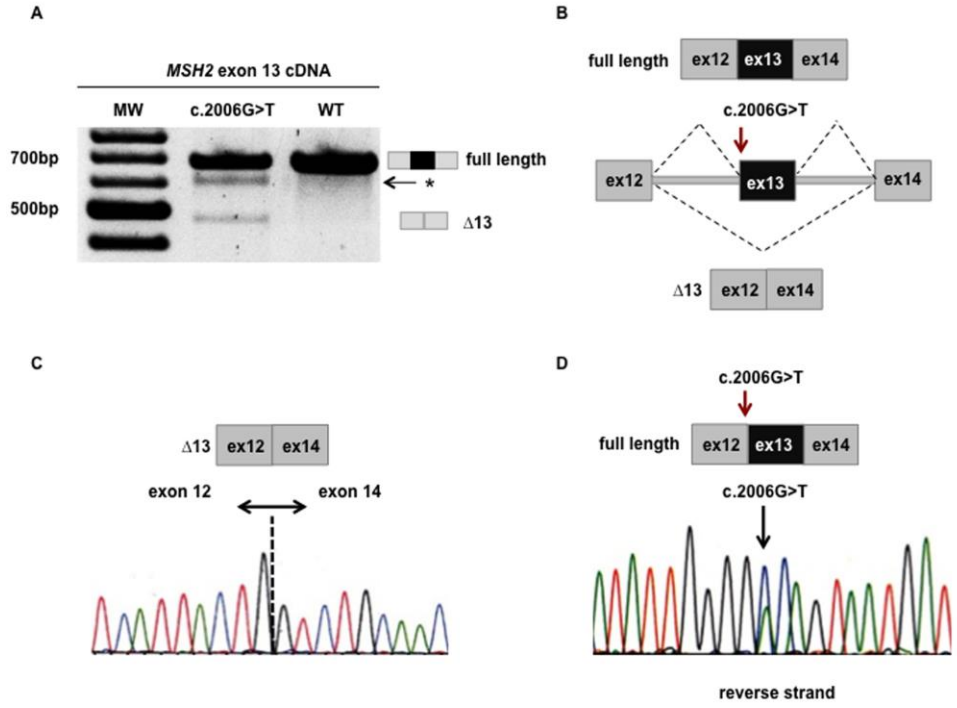
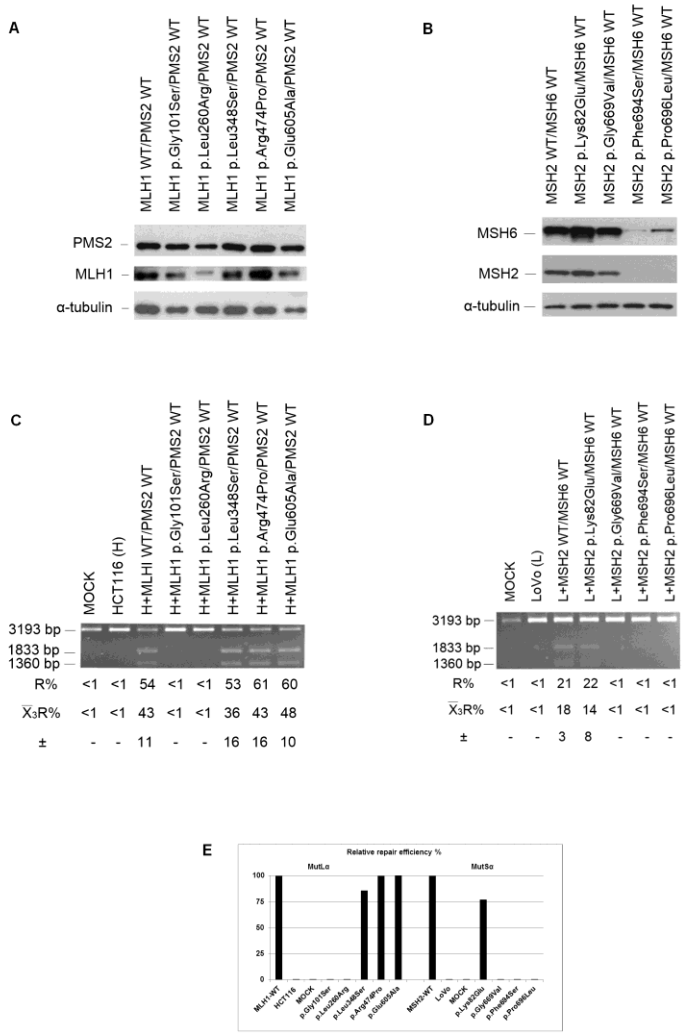


Figure 3. Expression and functional analyses of the 5 MLH1 and 4 MSH2 missense variants. (A-B) Western Blot analysis of total protein extracts from *Sf9* cells coinfected with baculovirus constructs expressing PMS2 wild-type protein (PMS2 WT) with either MLH1 WT or with MLH1 variant, and MSH6 WT with either MSH2 WT or MSH2 variant, and showing instability of proteins

MLH1 p.Leu260Arg, MSH2 p.Phe694Ser, and MSH2 p.Pro696Leu. α-tubulin was used as a loading control. (C-D) Repair efficiency (R%) of the recombinant MutLα (MLH1+PMS2) and Mutsα (MSH2+MSH6) protein variants measured in the *in vitro* MMR assay and calculated as the ratio of double digested DNA relative to total DNA added to the reaction, and showing functional deficiency in MLH1 p.Gly101Ser, MLH1 p.Leu260Arg, MSH2 p.Gly669Val, MSH2 p.Phe694Ser, and MSH2 p.Pro696Leu. R% corresponds to the assay results shown in the figure, $\bar{X}_3$R% denotes the average of three independent experiments with standard deviations (±). Nuclear extract free MOCK and uncomplemented MMR-deficient HCT116 NE and LoVo NE serve as negative controls and HCT116 and LoVo NE complemented by MutLα and Mutsα, respectively, serve as positive controls. The top fragment (3193 bp) shows the migration of unrepaired linearized G·T mismatch containing construct and the two smaller fragments (1833 bp and 1360 bp) represent the repaired and double digested fragments. (E) The relative repair % was calculated in respect to the wild type control (MutLα or MutSα respectively) set as 100 (100, $\bar{X}_3$R%) according to Drost et al., 2010, and Kantelinen et al., 2012.



931

932 Tabel 1.Population frequency and tumor molecular and immunohistochemical characteristics of MMR gene variants.

	Sequence variant			Population frequency		Tumor data				
Gene	Location	Nucleotide change[1]	Predicted aminoacid change[1]	ExAC[2,3]	Italian control chromosomes[4]	MSI status[6]	IHC[6]	*BRAF* p.Val600Glu[6]	*MLH1 promoter methylation*[6,7]	LOH[6]
ML	3	c.301G>	p.(Gly101	nr	0	MSI	nt	nt	wt	nt

H1		A	Ser)			-H (1)			
	3i	c.307-19A>G		0.00020	0	MSS (1)	normal (1)	nt	nt	no (1)
	9	c.779T>G	p.(Leu260Arg)	nr	0	MSI-H (6); MSS (1)	MLH1 loss (5: all MSI-H tumors)	wt (3)	wt (3: all with MLH1 loss)	no (3)
	11i	c.1039-8T>A		0.00155	0.021875	MSI-H (1)	MLH1 loss (1: tumor with MSI status unknown); MSH2 loss (1: MSI-H tumor)	nt	nt	nt
	12	c.1043T>C	p.(Leu348Ser)	0.00002	0	MSS (1)	normal (1)	nt	nt	no (1)
	12	c.1217G>A	p.(Ser406Asn)	0.00089	0.00333	MSI-H (2)	MLH1 loss (1); MSH2/MSH6 loss (1)	nt	wt (1: tumor with MLH1 loss)	variant allele (1: tumor with MLH1 loss)

	13	c.1421G>C	p.(Arg474Pro)	nr	0	MSI-H	MLH1 loss	mut	nt	variant allele
	13i	c.1558+14G>A		0.03948	0.021876	MSI-H (1)	MLH1 loss (1: tumor with MSI status unknown); MSH2 loss(1: MSI-H tumor)	nt	nt	nt
	15i	c.1732-19T>C		nr	0	MSS (1)	normal	nt	nt	variant allele
	16	c.1743G>A	p.(Pro581=)	0.00008	0	MSI-H (1)	MLH1 loss	mut	wt	nt
	16	c.1814A>C	p.(Glu605Ala)	nr	0	MSI-H (1)	MLH1 loss	nt	nt	no
	18	c.2041G>A	p.(Ala681Thr)	nr	0	MSS (1); MSI-H (1)	MLH1 loss (1: MSS tumor); normal (1: MSI-H tumor)	wt (1: MSI-H tumor)	wt (1: MSI-H tumor)	no (1: MSI-H tumor)
	18	c.2059C>T	p.(Arg687Trp)	0.00003	0	MSI-H (2); MSS	normal (1: MSI-H tumor); MLH1	wt (1: MSI-H tumor)	nt	nt

						(1)	loss (1: MSS tumor)			
MSH2	2	c.244A>G	p.(Lys82Glu)	nr	0	MSS (1)	normal	nt	nt	no (1)
	3	c.380A>G	p.(Asn127Ser)	0.00692	nt	nt	nt	nt	nt	nt
	81	c.1387-8G>T		0.00194	nt	MSI-L (1); MSS (1)	normal (1: MSI-L tumor); MLH1 loss (1: MSS tumor)	nt	nt	no (1: MSS tumor)
	9i	c.1511-9A>T		0.08400	0.07333	MSI-H (5); MSS (2)	normal (3: 1 MSI-H and 2 MSS tumors); MSH2/MSH6 loss (4: all MSI-H tumors)	nt	meth (1: MSI-H tumor)	nt
	11	c.1666T>C	p.(Pro556=)	0.00437	0	MSI-H (1); MSS (1)	normal (2)	nt	wt (1: MSI-H tumor)	nt

	11	c.1737A>G	p.(Lys579=)	0.00192	0	MSI-L (1); MSS (1)	normal (1: MSI-L tumor); MLH1 loss (1: MSS tumor)	nt	nt	no (1: MSS tumor)
	12i	c.2006-6T>C		0.11500	nt	MSI-H (2)	MSH2/MSH6 loss (2)	nt	nt	nt
	13	c.2006G>T	reported as p.(Pro670Leufs*) (predicted missense change: p.(Gly669Val)	nr	0	MSI-H (1)	MSH2/MSH6 loss (1)	nt	nt	no (1)
	13	c.2081T>C	p.(Phe694Ser)	nr	0	MSI-H (3)	MSH2/MSH6 loss (3)	nt	nt	nt
	13	c.2087C>T	p.(Pro696Leu)	nr	0	MSI-H (3)	MSH2/MSH6 loss (1)	nt	nt	no (1)
	14	c.2442T>G	p.(Leu814=)	nr	nt	MSI-H (1)	MSH2/MSH6 loss (1)	nt	nt	no (1)

933

934 [1] Previously unclassified variants are indicated in bold.

935 [2] nr = not reported.

936 [3] ExAC: http://exac.broadinstitute.org/; TGCA allele frequencies are excluded.

937 [4] nt = not tested.

938 [5] c.1039-8T>A and c.1558+14G>A are in linkage disequilibrium in the Italian population.

939 [6] In brackets number of samples; nt = not tested.

940 [7] wt = not methylated, meth = methylated.

941

942

943

944 **Table 2.** Effects of MMR gene variants on RNA processing, protein stability and *in vitro* MMR activity.

945

Gene	Sequence variant			mRNA analysis[2,3]			Functional analysis[3,4]	
	Location	Nucleotide change[1]	Predicted aminoacid change[1]	Studies on patient samples		Minigene splicing assay[4]	Mammalian protein stability	Mammalian mmr activity
				Splicing analysis	SNUPE assay			
MLH1	3	c.301G>A	p.(Gly101Ser)	nt	no allelic imbalance	total inclusion of reference exon 3 (Tournier et al., 2008; this study) with concomitant loss of alternative 5'ss five nucleotides upstream the reference5' ss	Stable	Deficient

						(this study)		
	3i	c.307-19A>G		nt	nt	no effect (Tournier et al. 2008)	na	na
	9	c.779T>G	p.(Leu260Arg)	no effect (Montera et al., 2000; this study)	no allelic imbalance	no effect	Unstable	Deficient
	11i	c.1039-8T>A		partial loss of exon 12	nt	no effect (Petersen et al., 2013; this study)	na	na
	12	c.1043T>C	p.(Leu348Ser)	no effect	no allelic imbalance	no effect	Stable	Proficient
	12	c.1217G>A	p.(Ser406Asn)	no effect	no allelic imbalance (Pastrello et al. 2011 and this study)	no effect	Stable (Takahashi et al. 2007)	Proficient (Takahashi et al. 2007; Drost et al. 2010)
	13	c.1421G>C	p.(Arg474Pro)	nt	nt	no effect	Stable	Proficient
	13i	c.1558+14G>A		nt	nt	no effect	na	na
	15i	c.1732-		no effect	nt	no effect	na	na

		19T>C						
	16	c.1743G>A	p.(Pro581=)	nt	nt	no effect	na	na
	16	c.1814A>C	p.(Glu605Ala)	nt	no allelic imbalance	no effect	Stable	Proficient
	18	c.2041G>A	p.(Ala681Thr)	no effect (Jakubowska et al., 2001; Betz et al., 2010)	no allelic imbalance	no effect (Tournier et al., 2008)	Discordant results (Raevaara et al., 2005; Takahashi et al., 2007; Xie et al., 2010; Hardt et al., 2011; Hinrichsen et al., 2013)	Proficient (Raevaara et al., 2005; Takahashi et al., 2007; Hinrichsen et al., 2013)
	18	c.2059C>T	p.(Arg687Trp)	no effect (Jakubowska et al., 2001; Furukawa et al., 2002; Auclair et al., 2006; Arnold et al., 2009; this study)	no allelic imbalance	no effect	Discordant results (Takahashi et al., 2007; Christensen et al., 2009)	Proficient (Takahashi et al., 2007; Christensen et al., 2009)
MS	2	c. 244A>G	p.(Lys82Glu	nt	no	no effect	Stable	Proficie

H2					allelic imbalance			nt
	3	c.380A>G	p.(Asn127Ser)	nt	nt	partial exclusion of exon 3	Stable (Kansikas et al., 2011)	Proficient (Ollila et al., 2008)
	8i	c.1387-8G>T		nt	nt	no effect (Tournier et al., 2008)	na	na
	9i	c.1511-9A>T		nt	nt	no effect (Tournier et al., 2008)	na	na
	11	c.1666T>C	p.(Pro556=)	no effect (Auclair et al., 2006)	nt	no effect (Tournier et al., 2008)	na	na
	11	c.1737A>G	p.(Lys579=)	no effect (Auclair et al., 2006)	nt	no effect (Tournier et al., 2008)	na	na
	12i	c.2006-6T>C5		no effect (Tournier et al., 2008)	na	partial exon 13 exclusion (Tournier et al., 2008; this study)	na	na
	13	c.2006G>T	reported as p.(Pro670Leufs*) (predicted missense	partial exon 13 exclusion (this study)	nt	complete exon 13 exclusion (van der Klift	Stable	Deficient

			change p.(Gly669Val)	complete exon 13 exclusion (van der Klift et al., 2015)		et al., 2015; this study)		
	13	c.2081T>C [5]	p.(Phe694Ser)	nt	no allelic imbalance	no effect	Unstable	Deficient
	13	c.2087C>T	p.(Pro696Leu)	nt	no allellic imbalance	no effect (Tournier et al., 2008)	Unstable	Deficient
	14	c.2442T>G	p.(Leu814=)	nt	nt	no effect	na	na

[1] Previously unclassified variants are shown in bold. [2] nt = not tested. [3] The results shown are from this study, unless otherwise indicated. [4] In italics: results of studies using *in vitro* mammalian assays different from that used in the present study. [5] These two variants (*MSH2* c.2006-6T>C and c.2081T>C) were also tested in combination in the minigene assay, since they were found in linkage disequilibrium.

Table 3. Clinical classification of *MLH1* and *MSH2* variants.

GENE	DNA and predicted protein change[1]	InSiGHT classification[2,3]	Proposed classification	Rationale for classification[4]	
				Posterior probability of pathogenicity by multifactorial analysis[3]	Qualitative criteria
MLH1	c.301G>A; p.(Gly101Ser)	4	5	0.99740	– Co-segregation – MSI/IHC data – Deficient MMR

					function – Allelic frequency: 0 (this study)
	c.307-19A>G	2	2	na	– Intronic location – Normal minigene splicing assay – MSI/IHC data – Allelic frequency: 0.0002 (ExAc); 0 (this study)
	c.779T>G; p.(Leu260Arg)	5	5	1	– Co-segregation – MSI/IHC data – Deficient MMR function – Allelic frequency: 0 (this study)
	c.1039-8T>A	1	1	na	– Intronic location – Allelic frequency: 0.00155 (ExAc); 0.02187 (this study) – MSI/IHC data – No major splicing abnormalities
	c.1043T>C; p.(Leu348Ser)		3	0.64637	Insufficient data (proficient MMR function; 1 MSS tumor; no major splicing alteration)
	c.1217G>A; p.(Ser406Asn)	1	1	< 0.00100	– Allele frequency: 0.00089 (ExAc);

					0.00333 (this study) – MSI/IHC data – Lack of co-segregation with phenotype (combined segregation likelihood ratio < 0.01) – Co-occurrence of MSH2 truncating variant that segregates with the phenotype in the family – Proficient MMR function – No major splicing abnormalities – Estimated risk from case-control studies (1.5)
	c.1421G>C; p.(Arg474Pro)	3	3	0.09448	Insufficient data (no major splicing alteration; proficient MMR function; 1 tumor MSI-H BRAF p.Val600Glu positive)
	c.1558+14G>A	1	1	na	– Intronic location – Allelic frequency: 0.03948 (ExAc); 0.02187 (this study) – MSI/IHC data

					– No major splicing abnormalities
	c.1732-19T>C	na	2	0.01386	– Intronic location – 1 MSS tumor – No major splicing abnormalities
	c.1743G>A; p.(Pro581=)	na	2	na	– Synonymous coding change – No major splicing abnormalities by minigene assay – Allelic frequency: 0.00008 (ExAc); 0 (this study) – 1 MSI-H MLH1-neg BRAF p.Val600Glu-pos tumor
	c.1814A>C; p.(Glu605Ala)	na	4	0.95294	MSI/IHC data
	c.2041G>A; p.(Ala681Thr)	5	5	0.99708	– Co-segregation – MSI/IHC data
	c.2059C>T; p.(Arg687Trp)	5	5	0.99999	– Co-segregation – MSI/IHC data – Homozygosity associated with constitutional mismatch repair deficiency syndrome
MSH 2	c. 244A>G; p.(Lys82Glu)	na	2	0.00980	– MSI/IHC data – Proficient MMR function

					– No major splicing abnormalities
	c.380A>G; p.(Asn127Ser)	1	1	na	– Allelic frequency: 0.0692 (ExAc) – No major splicing abnormalities by minigene assay – Proficient MMR function
	c.1387-8G>T	2	1	0.00088	– Intronic location – Allelic frequency: 0.00194 (ExAc) – MSI/IHC data (> 3 tumors not showing features of MMR deficiency and/or MSH2 inactivation) – No major splicing abnormalities
	c.1511-9A>T	1	1	na	– Intronic location – Allelic frequency: 0.08400 (ExAc); 0.07333 (this study) – MSI/IHC data – No major splicing abnormalities by minigene

					assay
	c.1666T>C; p.(Pro556=)	1	1	< 0.00010	– Synonymous coding change – Allele frequency: 0.00437 (ExAc); 0 (this study) – MSI/IHC data – No major splicing abnormalities
	c.1737A>G; p.(Lys579=)	2	1	0.00021	– Synonymous coding change – Allelic frequency: 0.0019 (ExAc); 0 (this study) – MSI/IHC data (> 3 tumors not showing features of MMR deficiency and/or MSH2 inactivation) – No major splicing abnormalities
	c.2006-6T>C	1	1	na	– Intronic location – No major splicing abnormalities
	c.2006G>T; reported as p.(Pro670Leufs*) (predicted missense change p.(Gly669Val)	5	5	0.99906	– Co-segregation – MSI/IHC data – Deficient MMR functional test – Contrasting results of RNA splicing analyses

	c.2081T>C; p.(Phe694Ser)	5	5	0.99990	– Co-segregation – MSI/IHC data – Deficient MMR function
	c.2087C>T; p.(Pro696Leu)	5	5	1	– Co-segregation – MSI/IHC data – Deficient MMR function
	c.2442T>G; p.(Leu814=)	na	2	na	– Synonymous coding change – No major splicing abnormalities by minigene splicing assay – Co-observation of MSH2 Class 5-pathogenic variant (phase unknown)

[1]Variants not yet classified by InSiGHT are shown in bold. [2] For previously classified variants, the classification corresponds to that reported on http://insight-group.org/variants/classifications/.

[3] na = not available. [4] Classification was achieved by multifactorial analysis, qualitative criteria or both; in italics data obtained at least in part from the present study.